ARL-STRUC-TM-371 AR-002-997 12

AD-A142 810

DEPARTMENT OF DEFENCE

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION

AERONAUTICAL RESEARCH LABORATORIES

MELBOURNE, VICTORIA

Structures Technical Memorandum 371

HOW MANY SPECIMENS? - AN AID IN THE DESIGN OF FATIGUE TEST PROGRAMS

A.S. MACHIN

Approved for Public Release

JANUARY 1984

UNCLASSIFIED

AR-002-997

DEPARTMENT OF DEFENCE

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION

AERONAUTICAL RESEARCH LABORATORIES

Structures Technical Memorandum 371

HOW MANY SPECIMENS? - AN AID IN THE DESIGN OF FATIGUE TEST PROGRAMS

by

A.S. Machin

SUMMARY

Increasing fatigue test sample sizes increases the accuracy with which population statistics may be calculated and increases the sensitivity of tests comparing different groups of results, but it also increases testing costs. Information is presentated which will aid the designer of a fatigue test program in deciding sample sizes.

POSTAL ADDRESS: Director, Aeronautical Research Laboratories,
P.O. Box 4331, Melbourne, Victoria, 3001, Australia.

CONTENTS

NOTATION

n	size of sample
s	sample standard deviation
S_c	weighted, combined standard deviation of two samples
$t_{m,\alpha}$	percentage point of Student's t-distribution with m degrees of freedom and significance α
$\bar{x}$	sample mean
$\chi^2_{m,\alpha}$	percentage point of chi-square distribution with m degrees of freedom and significance α
μ	population mean
σ	population standard deviation

1. INTRODUCTION

The purpose of this memorandum is to aid an experimenter with a portion of the design of Fatigue Test programs; the question of how many specimens to test at each condition. None of the information is novel or original, but it may be useful to those who are experimenters first and statisticians second. There is an extensive literature on the subject of Experimental Design (e.g. Refs. 1,2) which should be studied, in particular, those sections on randomisation in experimental design should be noted, since valid statistical inference depends on having a 'random sample'.

Fatigue testing can be expensive (Appendix 1). The overall cost of a Fatigue Test program can be reduced by minimising the number of specimens tested, but it is essential to test enough specimens so that reliable statistical inferences* may be extracted from the results. Given that resources are often limited it may be necessary to reduce the scope of a test program so that a restricted amount of reliable information is obtained rather than a quantity of potentially inconclusive data.

2. STATISTICAL ASSUMPTIONS

In the analysis of fatigue test data, a common practice is to assume that the logarithms (to base 10) of the fatigue lives of specimens are members of a population belonging to a normal distribution. Other theoretical frequency distributions are used (e.g. the various extreme-value distributions) but the information presented in this Memorandum is derived from analysis of the normal distribution. For distributions which are not quite normal, the results presented are adequate for most practical purposes (Ref. 3).

The sample statistics first calculated when analysing fatigue test data are the standard deviation (s) and the mean ($\bar{x}$), which is usually reported as antilog ($\bar{x}$). Once these have been calculated,

---

* Methods of statistical inference may be complex, in this Memorandum only three basic techniques will be examined.

much fatigue analysis involves comparing different groups of data (using a t-test) to see whether differences between sample means are significant.

The following sections show how the accuracy of estimation of population parameters such as the mean and standard deviation depend on the size of the sample and also how sample sizes influence the sensitivity of a t-test.

3. INTERVAL ESTIMATION OF MEANS

If $\bar{x}$ and s are the mean and standard deviation of a sample of size n taken from a normal distribution $N(\mu,\sigma^2)$ where μ and σ are unknown*, then:

$$\bar{x} - t_{n-1,\alpha/2} \frac{s}{\sqrt{n}} \leqslant \mu \leqslant \bar{x} + t_{n-1,\alpha/2} \frac{s}{\sqrt{n}} \tag{1}$$

defines a $100(1-\alpha)\%$ confidence interval for μ (Ref. 3).

Relation (1) may be modified to the form:

$$|\mu - \bar{x}| \leq t_{n-1,\alpha/2} \frac{s}{\sqrt{n}} \tag{2}$$

Since it is the logarithms of specimens lives which are assumed to belong to a normal distribution, it follows that $10^{|\mu-\bar{x}|}$ will give the ratio between the antilogs of the population mean μ and the sample mean, $\bar{x}$.

This information is presented in Figures 1 to 7 for sample standard deviations of 0.04, 0.06, 0.08, 0.10, 0.12, 0.16 and 0.20 respectively. In each Figure information is given for confidence intervals of 80%, 90%, 95% and 99%+.

---

* In fatigue test data analysis it is most unlikely that the population standard deviation will be known. If it is known, the interval estimation of the mean, with a small sample size, will be more precise.

+ For all Figures in this Memorandum the values of the relationships presented are calculated only at integer sample sizes and related points are connected by straight line segments. Do not interpolate between calculated points.

Figures 1 to 7 show that errors in estimation of the population mean, μ, (by using the sample mean $\bar{x}$) can be large. For a sample standard deviation of 0.10*; then, if the sample has three members, there is a 1 in 5 chance of the estimate of μ being in error by at least 28%, and a 1 in 20 chance of the estimate being in error by 77% (Fig. 4). Even with a large sample (by fatigue testing practice) there is a limit to the accuracy of estimation of μ. With 11 test results in a sample of standard deviation 0.10, there is a 1 in 5 chance of the estimate of μ being more than 10% in error and a 1 in 20 chance of the estimate being more than 17% in error.

The data presented in Figures 1 to 7 also show that improvements in accuracy of estimation of the population mean are dramatic between a sample size of three and a sample size of five or six but that increasing the sample size further yields much smaller improvements. Using again a sample of standard deviation 0.10; if there are three results in the sample there is a 1 in 20 chance of the estimate of the mean being more than 77% in error. If the testing cost is increased by doubling the sample size and six results obtained there is a 1 in 20 chance of the estimate of population mean being more than 27% in error. If the testing cost is again increased and the sample consists of 11 results there is a 1 in 20 chance of the estimate of the population mean being more than 17% in error.

4. INTERVAL ESTIMATION OF STANDARD DEVIATION

If s is the standard deviation of a sample of size n from a normal distribution (μ,σ^2) then:

$$\sqrt{\frac{n-1}{\chi^2_{n-1,\alpha/2}}} \leqslant \frac{\sigma}{s} \leqslant \sqrt{\frac{n-1}{\chi^2_{n-1,1-\alpha/2}}} \qquad (3)$$

defines a $100(1-\alpha)$% confidence interval for σ (Ref. 3).

---

* Standard deviations of about this magnitude are common if the testing is done in an electro-hydraulic fatigue testing machine (e.g. the results presented in Ref. 4) but the data presented in Ref. 5 suggests that larger standard deviations may expected.

Figure 8 presents the relation (3) for confidence intervals of 80%, 90%, 95% and 99%. This Figure shows the difficulty of making an accurate estimate of a population standard deviation, even with (by fatigue test practice) quite large sample sizes. For each confidence interval the degree by which the population standard deviation may be under-estimated (by using s) is greater than the degree to which it may be over-estimated.

As with estimation of population means, a sample size of five or six gives an estimate of population standard deviation which approaches the accuracy obtainable with a sample size of 11 and is much better than the accuracy obtained with a sample size of three.

5. THE t-TEST

This test may be used to establish whether there is a significant difference between two sample means*. If the means of the parent populations are μ_1, and μ_2 respectively and the hypothesis is made that $\mu_1 = \mu_2$, then this hypothesis may be tested by the relation (Ref. 3):

$$-t_{n_1+n_2-2,\ \alpha/2} \leq \frac{\bar{x}_1 - \bar{x}_2}{S_c(1/n_1 + 1/n_2)^{\frac{1}{2}}} \leq t_{n_1+n_2-2,\ \alpha/2} \qquad (4)$$

where S_c is the combined sample standard deviation and is given by:

$$S_c = \left(\frac{(n_1-1)S_1^2 + (n_1-1)S_2^2}{n_1 + n_2 - 2}\right)^{\frac{1}{2}}$$

and where α is the probability of rejecting the hypothesis when it is correct+.

---

* Sample variances should be compared first, using an F-test, to check that population variances are not significantly different.

\+ Rejecting a correct hypothesis is a Type 1 error. Users of the t-test should bear in mind the possibility of Type II errors, i.e., accepting as correct a hypothesis that is incorrect. The level of Type II errors depends on the deviation between the two populations, which in practice can not be determined.

If $\dfrac{\bar{x}_1 - \bar{x}_2}{S_c(1/n_1 + 1/n_2)^{\frac{1}{2}}}$ falls outside the limits given in (4) above, then the hypothesis that $\mu_1 = \mu_2$ is rejected and there is a significant difference between the sample means.

The relation (4) may be re-arranged to:

$$|\bar{x}_1 - \bar{x}_2| \leqslant t_{n_1 + n_2 - 2,\alpha/2}\, S_c\, (1/n_1 + 1/n_2)^{\frac{1}{2}} \tag{5}$$

Since logarithms of specimen lives are used, it follows that $10^{|\bar{x}_1 - \bar{x}_2|}$ will give the ratios between the antilogs of the two sample means.

Relation (5) is presented in Figures 9, 10, 11 for confidence levels of 10%, 5% and 1% respectively and for combined sample standard deviations of 0.06, 0.10, 0.12 and 0.16. The simplifying condition that $n_1 = n_2$ has been made.

These Figures show the effect on the sensitivity of a t-test of the sample size and combined sample standard deviation, and enable an estimate to be made of the differences in mean life that must exist between samples before the t-test will show them to be significantly different. Using Fig. 10 ($\alpha = 0.05$) as an example; if the combined sample deviation is 0.10 then, for samples of six results each, a life difference of 20% will not be significant but a life difference of 25% will be significant.

Even with large sample sizes, the differences in mean lives have to be considerable before it can be shown that they are significant. For example, for $\alpha = 0.05$, a combined sample standard deviation of 0.10 (Fig. 10) and two samples of 11 results each, the mean lives of the samples must differ by more than 16% for them to be significantly different.

Samples of five or six results each should provide a good compromise between fatigue test costs and t-test sensitivity.

6. CONCLUDING REMARKS

In a fatigue test program sample means or standard deviations cannot be nominated before testing (though a literature search of the results of similar tests may give a guide to the expected magnitude of standard deviations). This being the case it is impossible to examine the Figures in this Memorandum and so decide in advance the numbers of specimens to be tested to provide a desired accuracy of statistical inference.

The information presented here should provide the reader with some appreciation of the limits of accuracy of estimation of population parameters when using the small sample sizes that are usually forced by constraints on test program cost and time. It has been shown that, accuracy improves dramatically as the sample size is increased to five or six results, but improvement is less when sample sizes are increased further. It is up to the experimenter to decide when the costs of increasing sample size outweight the benefits of improved accuracy.

REFERENCES

1. Fisher, R.A. The design of experiments. Edinburgh, Oliver and Boyd, 1937.

2. Little, R.E. and Jebe, E.H. Statistical design of fatigue experiments. London, Applied Science Publishers Ltd. 1975.

3. Guttman, I. and Wilks, S.S. Introductory engineering statistics. New York, John Wiley, Inc. 1965. (See pp. 140-145, 168-169).

4. Mann, J.Y., Lupson, W.F., Machin, A.S. and Pell, R.A. Interim report on investigation to improve the fatigue life of the Mirage IIIO wing spar. Aero. Res. Labs. Structures Tech. Memo. 334, August 1981.

5. Mann, J.Y. Scatter in fatigue - a materials, testing and design problem. Materials, experimentation and design in fatigue. Proc. of Fatigue '81, 1981, pp 390-423.

APPENDIX I

COST OF TESTING SPECIMENS

A test program designed to improve the fatigue life of the main spar of the Mirage IIIO wing spar was recently completed by Structures Division of Aeronautical Research Laboratories. During this program 246 moderately complex specimens were tested in about 2.5 years, i.e. a testing rate of 100 specimens per year. A very simple estimate of the direct cost is given below.

COSTS

(1)	Material		Supplied by RAAF, no charge to A.R.L.
(2)	Machining of specimen blanks		\$250 per specimen
(3)	Cost of using testing machine	(a)	\$200 per specimen
(4)	Salaries	(b)	\$450 per specimen
		TOTAL	\$900 per specimen

(a) Assuming a machine life of 10 years and a capital cost of \$100,000 (1983 dollars) for an electrohydraulic fatigue testing machine and assuming maintenance and running costs of \$100,000 over 10 years, then cost per year will be \$20,000. If 100 specimens are tested per year, the cost per specimen will be \$200.

(b) Assuming direct salary costs of \$45,000 per year, i.e. approximately one sub-professional and one professional officer.

In addition to the costs above, various items of equipment were purchased as part of the program and many man-months were spent writing the final Report.

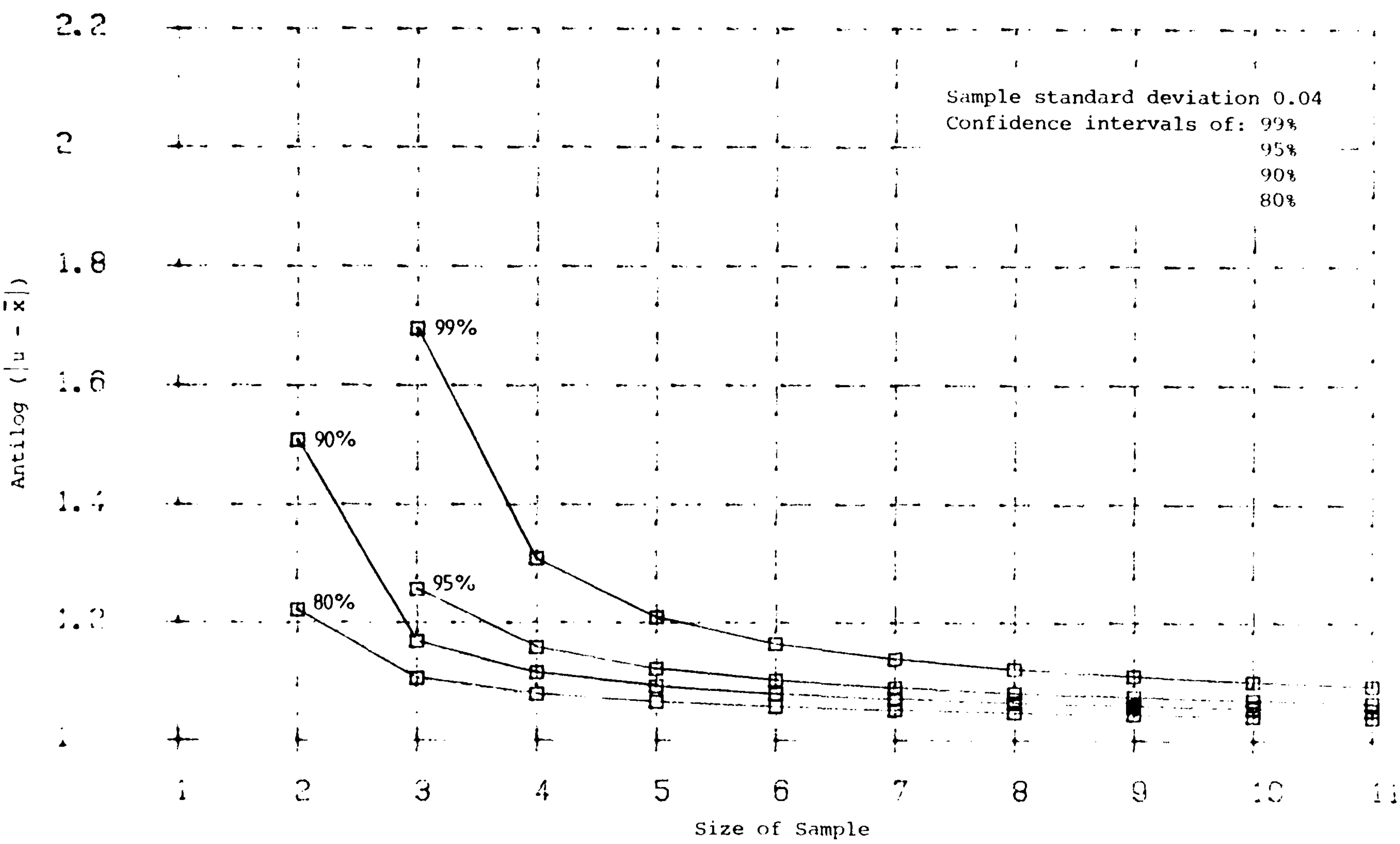

FIG. 1 CONFIDENCE INTERVALS FOR MEANS OF SAMPLES WITH A STANDARD DEVIATION OF 0.04. POPULATION STANDARD DEVIATION UNKNOWN.

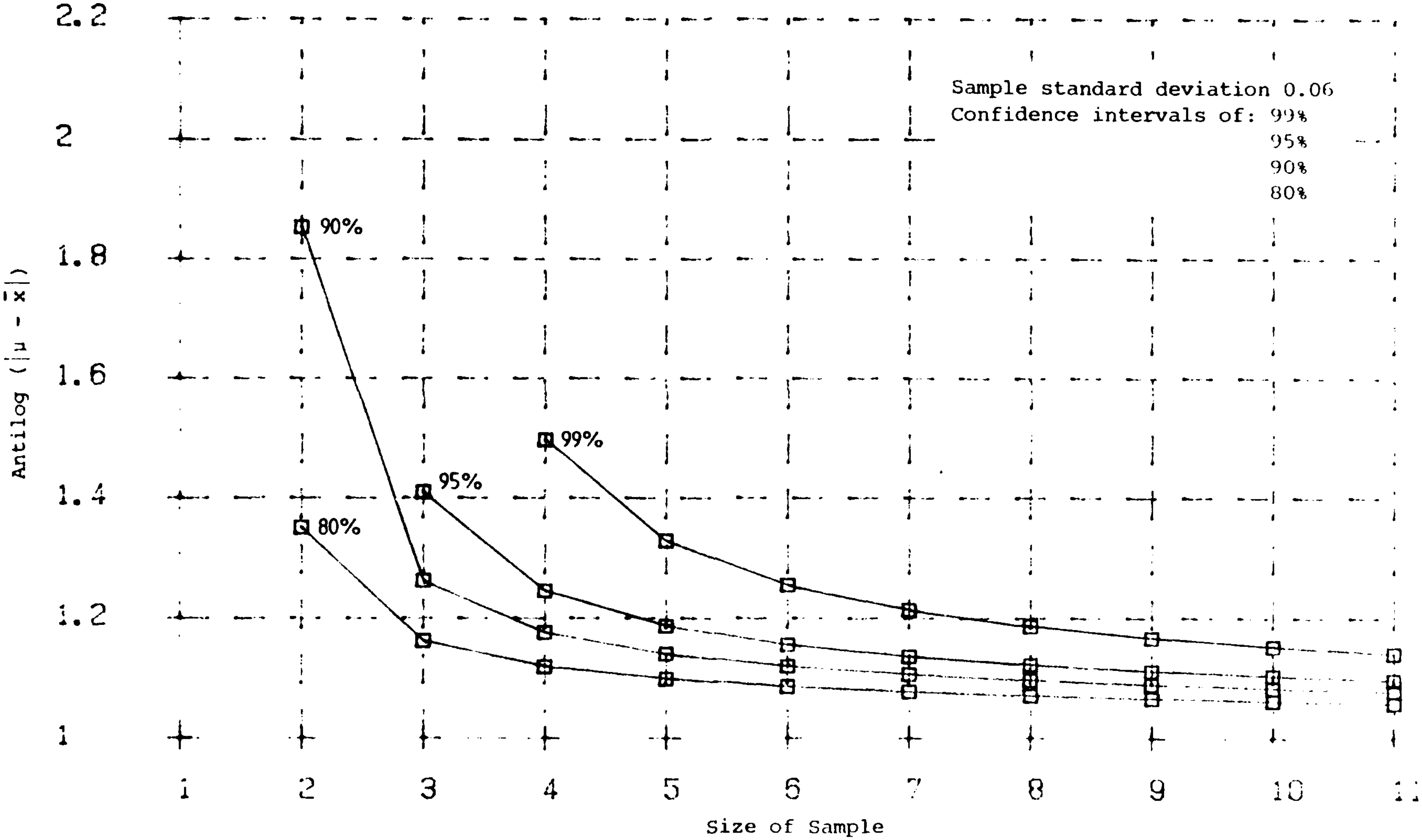

FIG. 2 CONFIDENCE INTERVALS FOR MEANS OF SAMPLES WITH A STANDARD DEVIATION OF 0.06. POPULATION STANDARD DEVIATION UNKOWN.

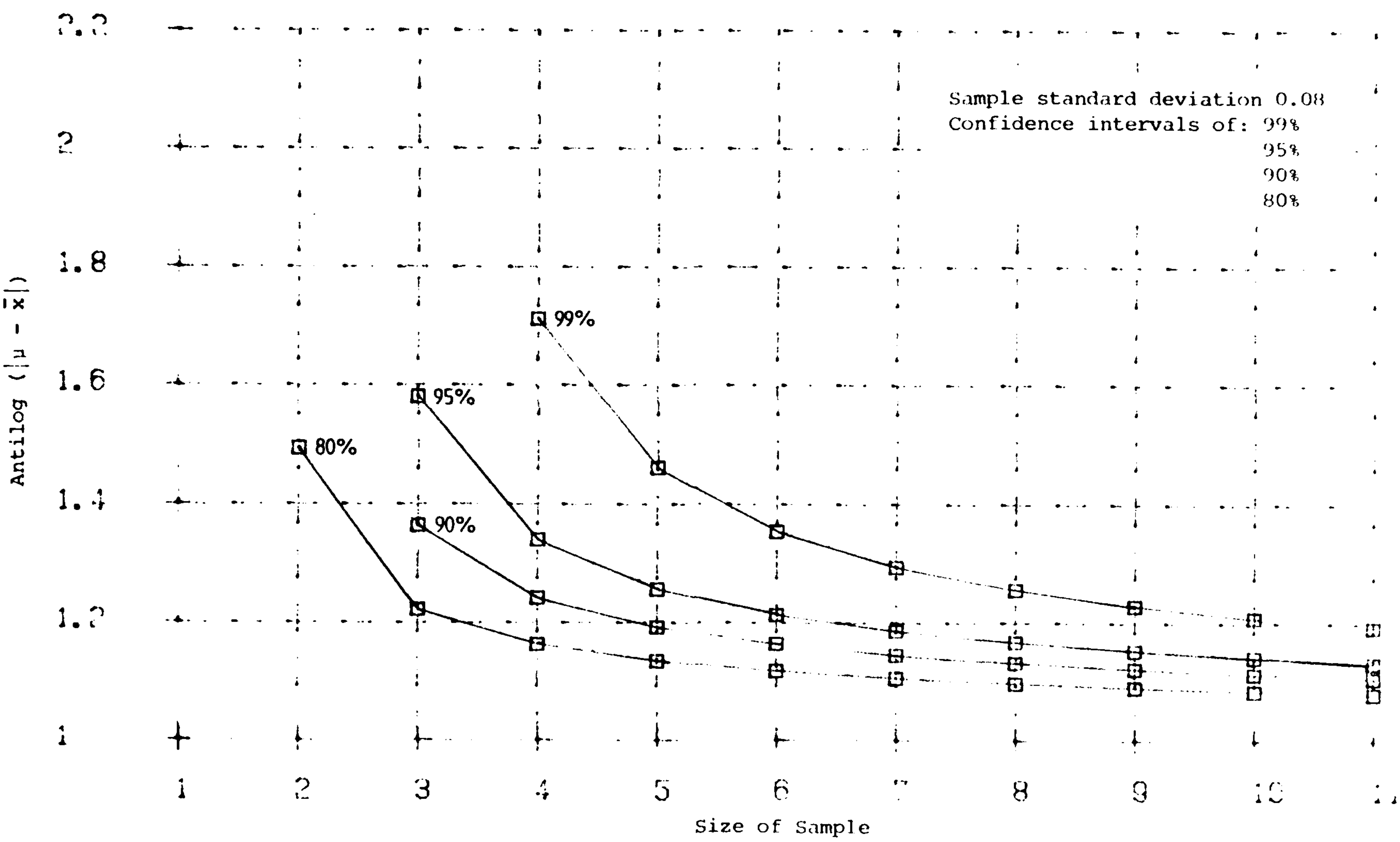

FIG. 3 CONFIDENCE INTERVALS FOR MEANS OF SAMPLES WITH A STANDARD DEVIATION OF 0.08. POPULATION STANDARD DEVIATION UNKNOWN.

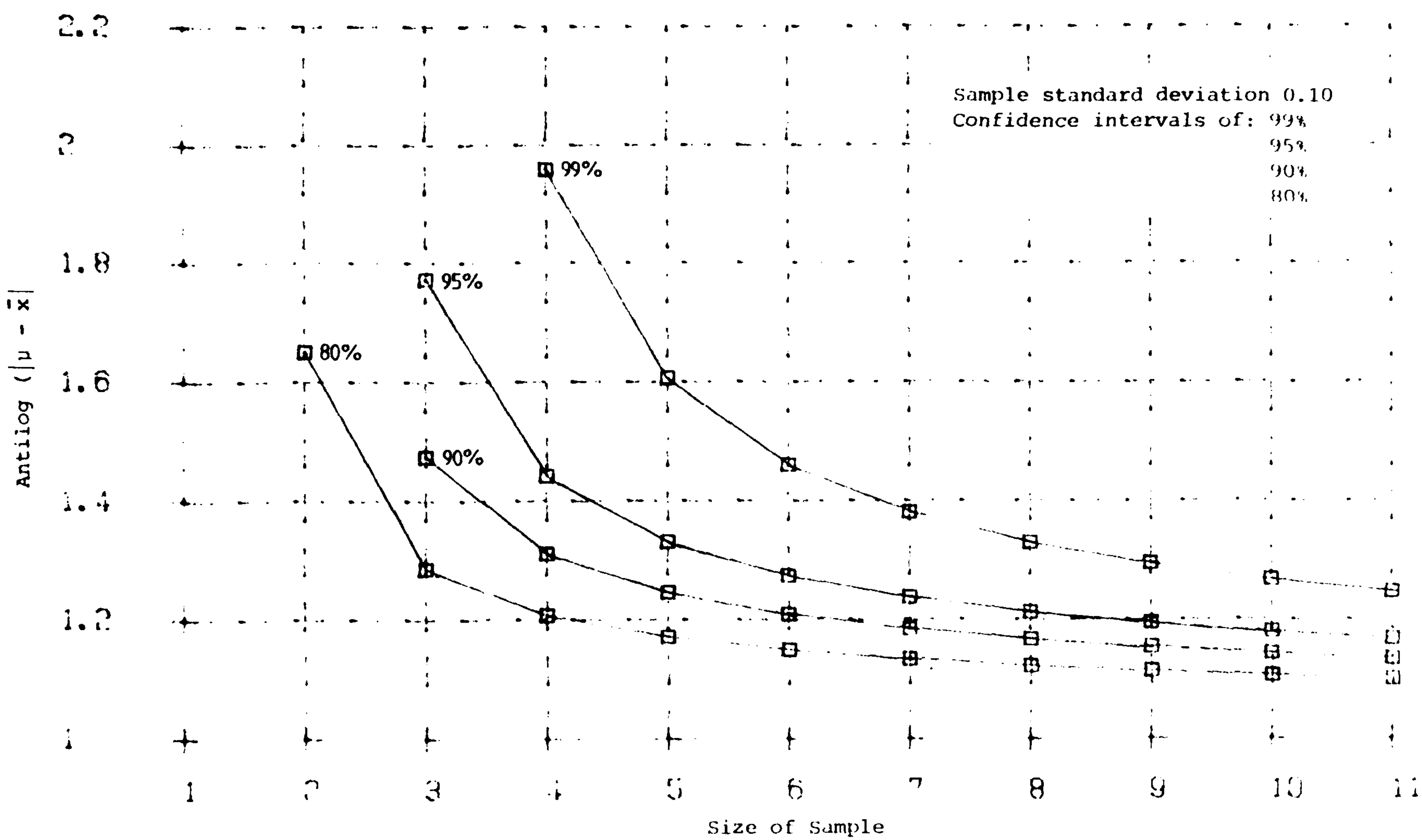

FIG. 4 CONFIDENCE INTERVALS FOR MEANS OF SAMPLES WITH A STANDARD DEVIATION OF 0.10. POPULATION STANDARD DEVIATION UNKNOWN.

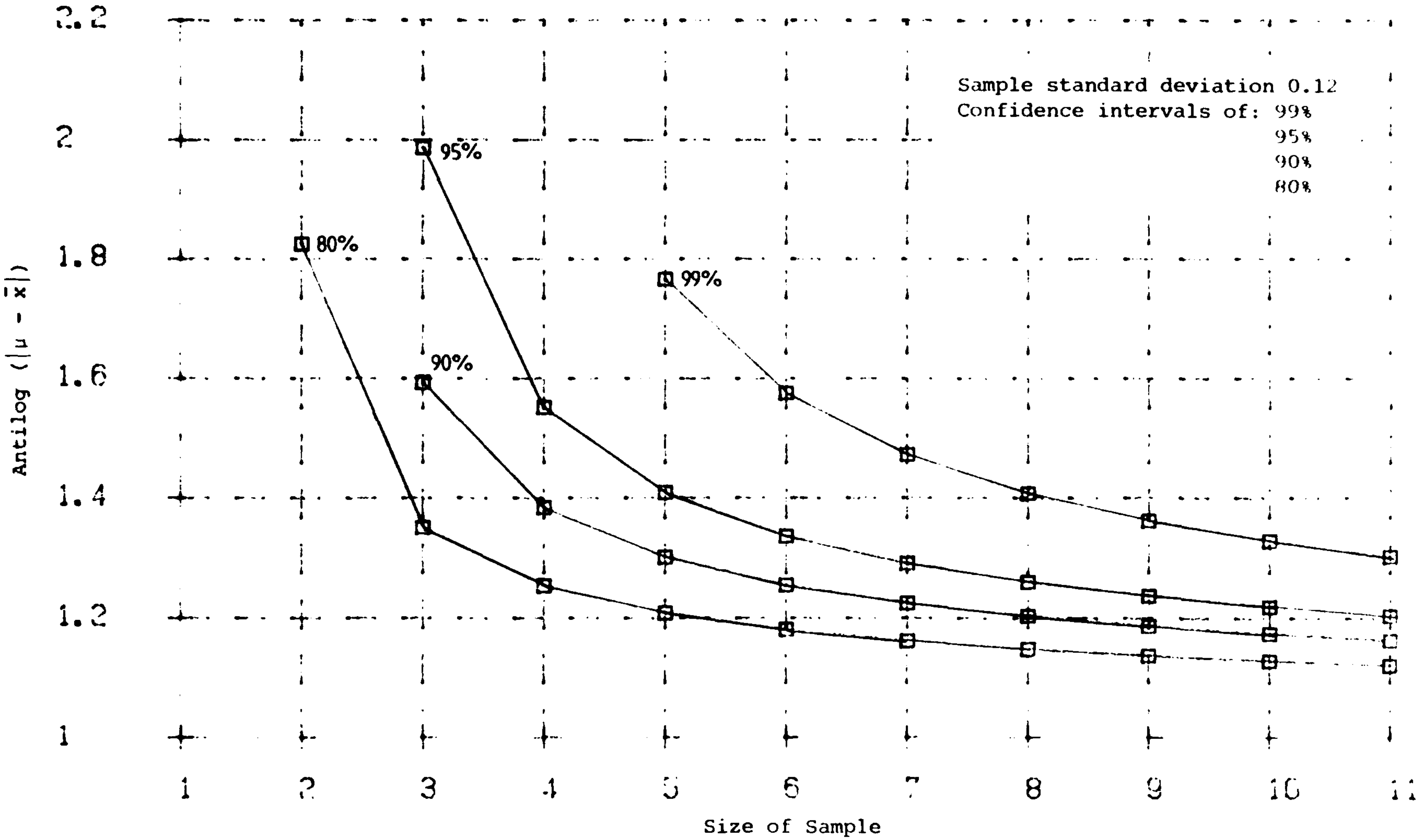

FIG. 5 CONFIDENCE INTERVALS FOR MEANS OF SAMPLES WITH A STANDARD DEVIATION OF 0.12. POPULATION STANDARD DEVIATION UNKNOWN.

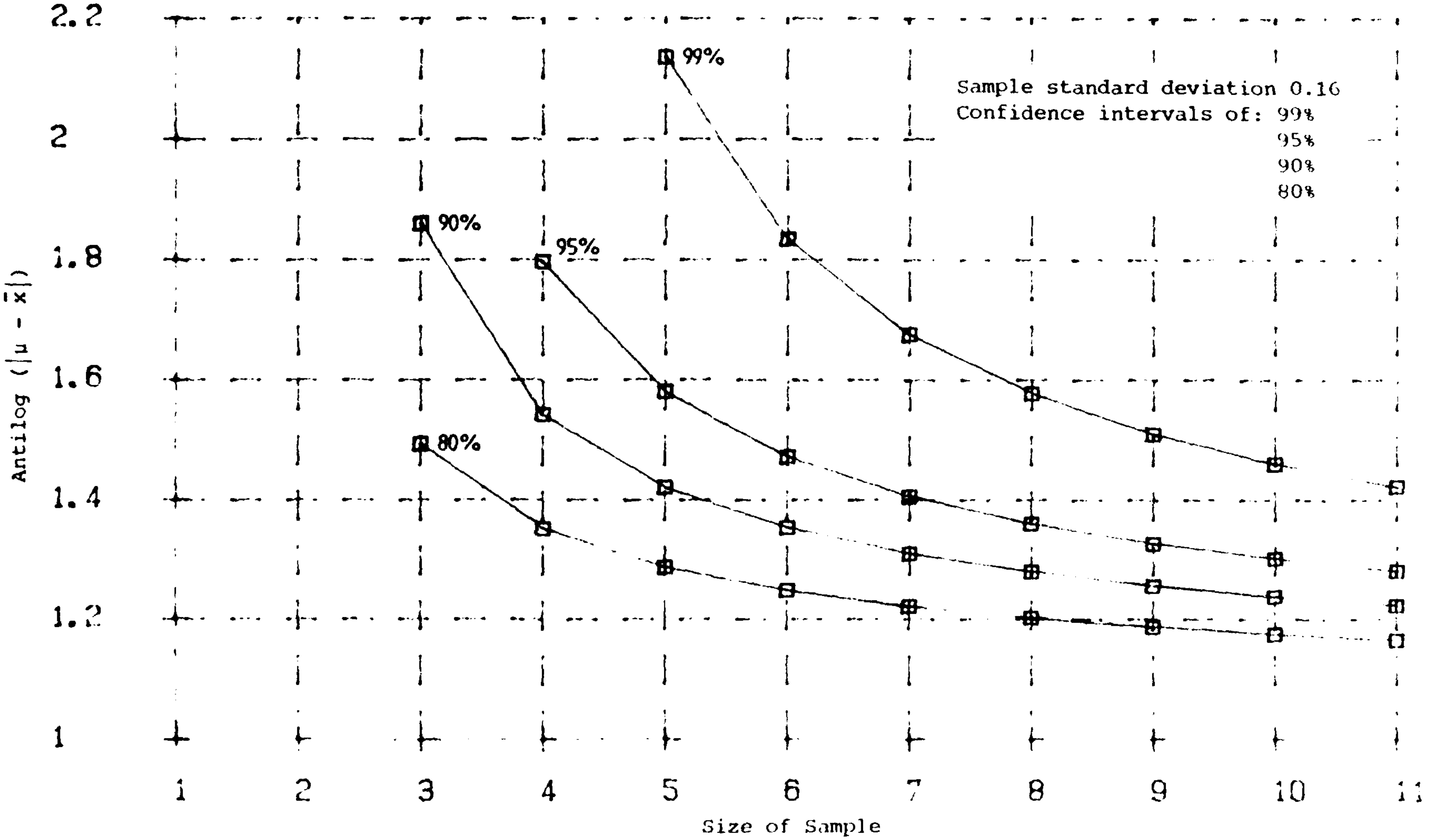

FIG. 6 CONFIDENCE INTERVALS FOR MEANS OF SAMPLES WITH A STANDARD DEVIATION OF 0.16. POPULATION STANDARD DEVIATION UNKNOWN.

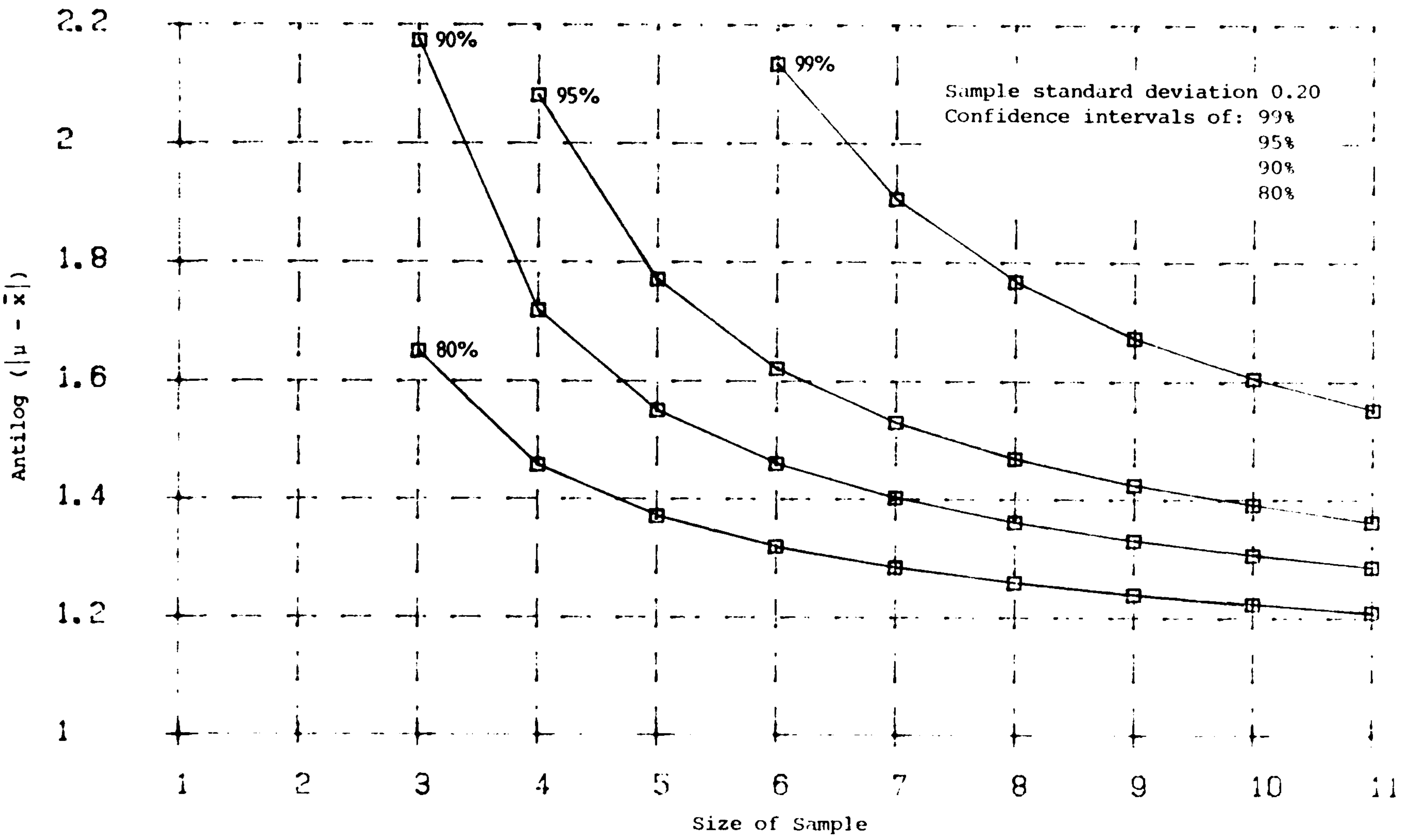

FIG. 7 CONFIDENCE INTERVALS FOR MEANS OF SAMPLES WITH A STANDARD DEVIATION OF 0.20. POPULATION STANDARD DEVIATION UNKNOWN.

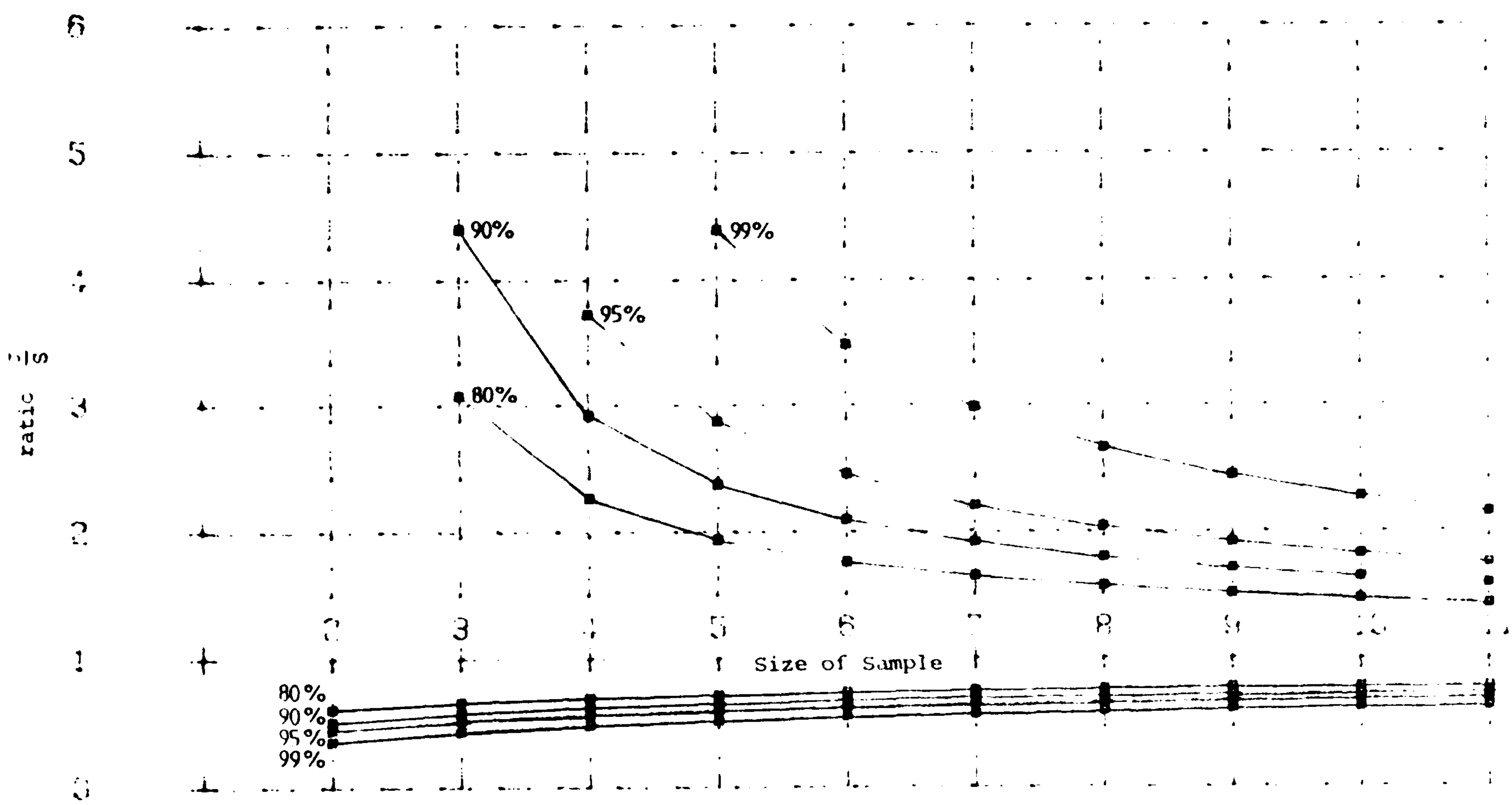

FIG.8 RATIO $\frac{\sigma}{s}$ FOR VARIOUS SAMPLE SIZES AND CONFIDENCE INTERVALS OF 80%, 90%, 95% AND 99%.

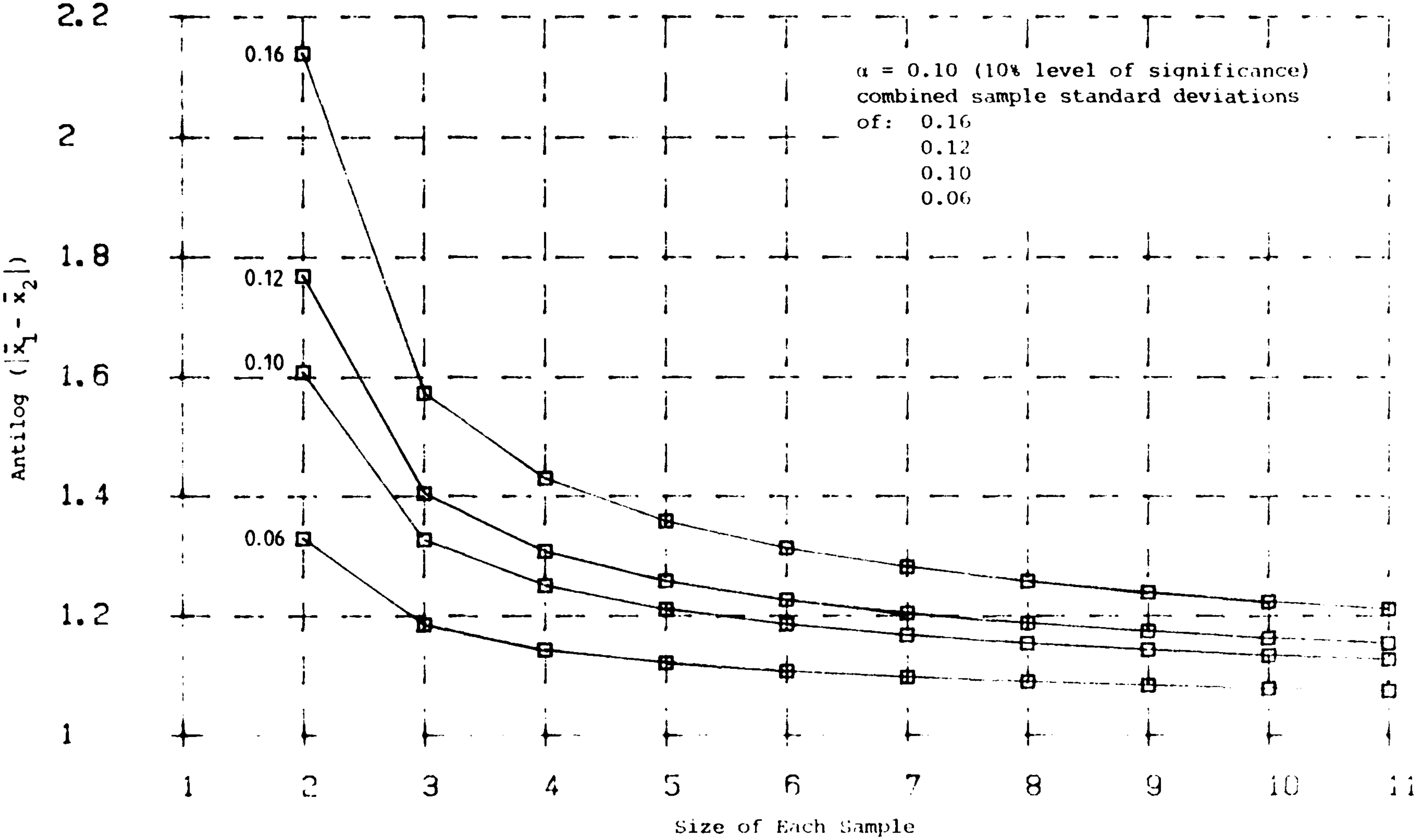

FIG. 9 EFFECT OF SAMPLE SIZE ON t-TEST SENSITIVITY AT THE 10% LEVEL OF SIGNIFICANCE.

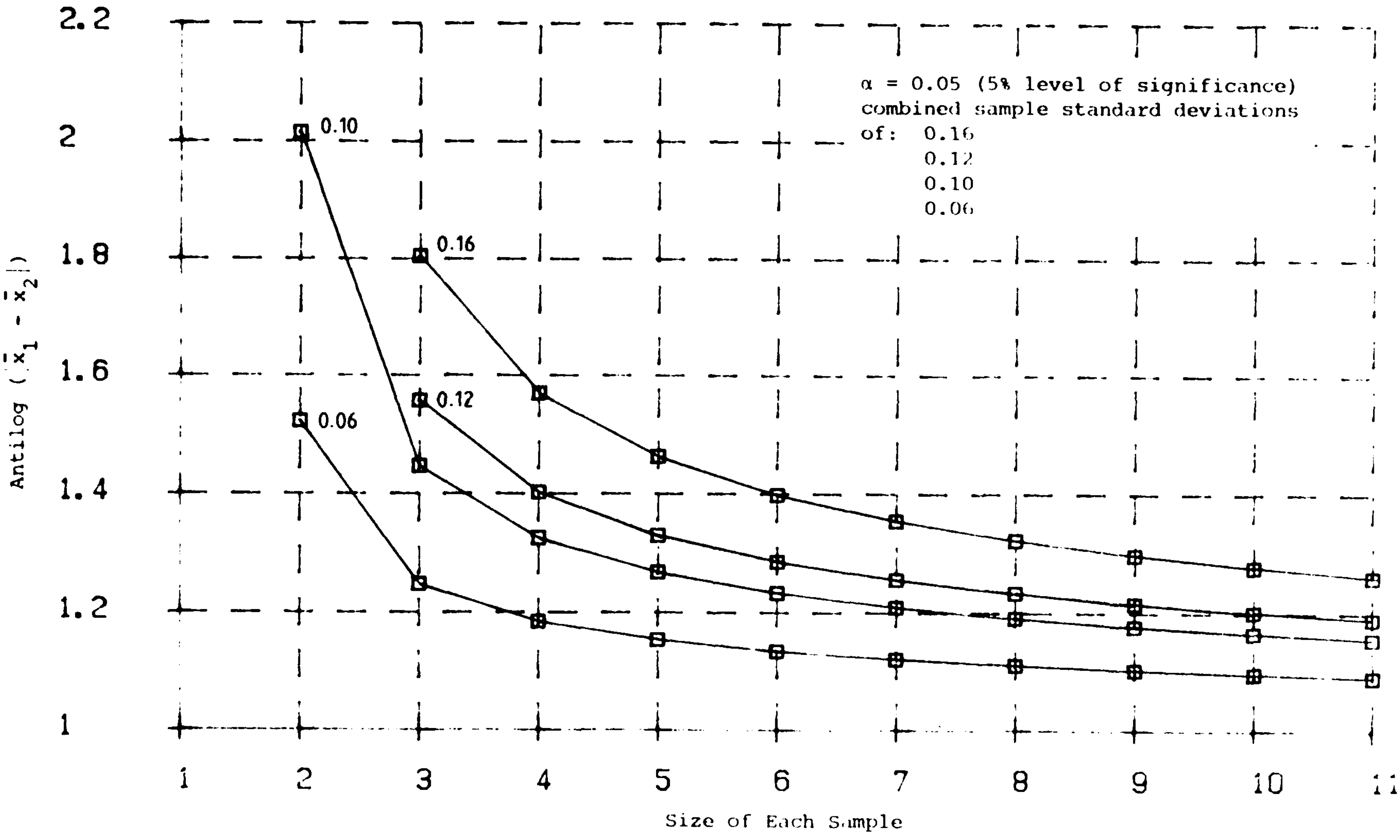

FIG. 10 EFFECT OF SAMPLE SIZE ON t-TEST SENSITIVITY AT THE 5% LEVEL OF SIGNIFICANCE.

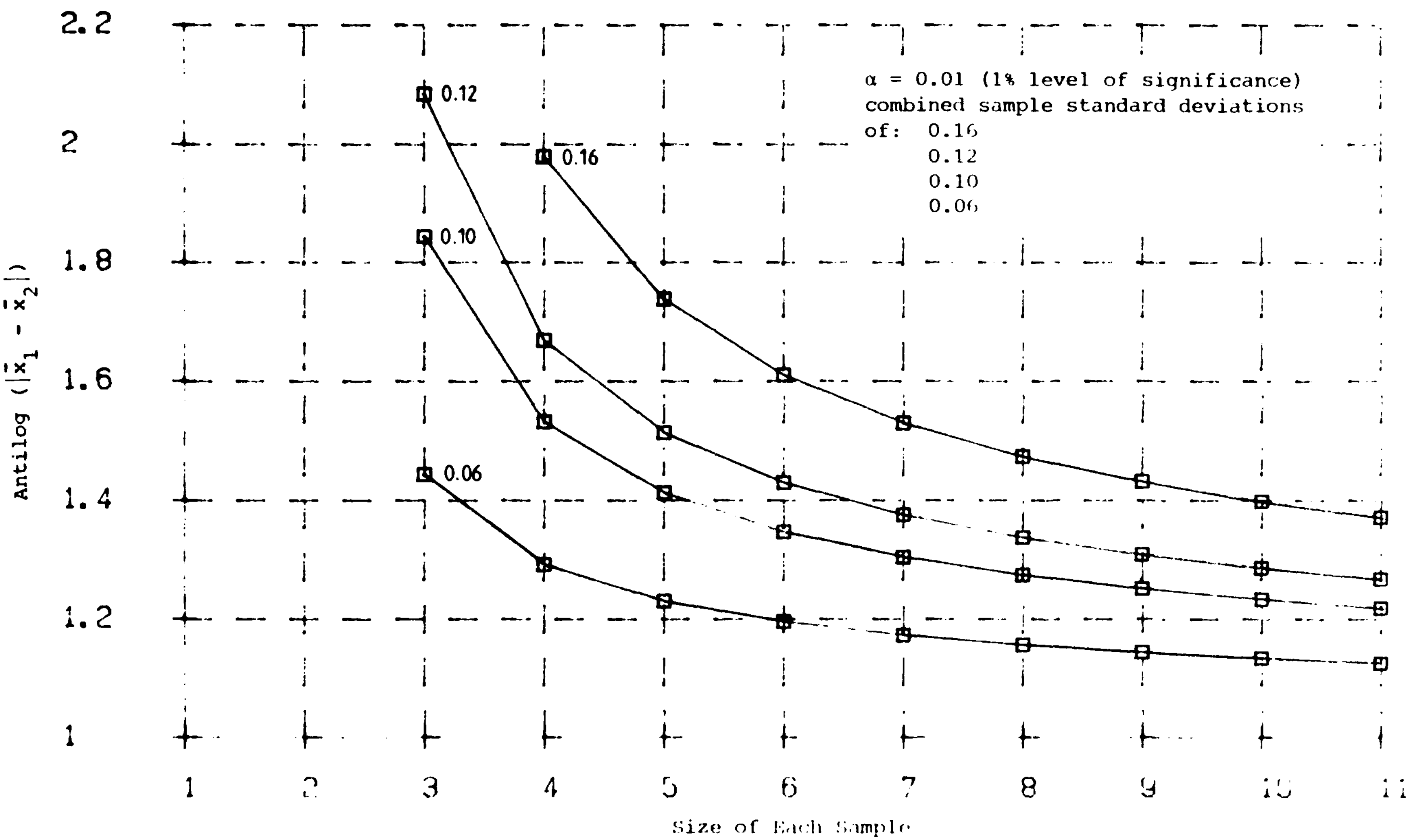

FIG. 11 EFFECT OF SAMPLE SIZE ON t-TEST SENSITIVITY AT THE 1% LEVEL OF SIGNIFICANCE.

DISTRIBUTION

AUSTRALIA

Department of Defence

Central Office

Chief Defence Scientist)
Deputy Chief Defence Scientist)
Superintendent, Science and Technology Programmes) (1 copy)
Controller, Projects and Analytical Studies)
Defence Science Representative (U.K.) (Doc Data sheet only)
Counsellor, Defence Science (U.S.A.) (Doc Data sheet only)
Defence Central Library
Document Exchange Centre, D.I.S.B. (18 copies)
Joint Intelligence Organisation
Librarian H Block, Victoria Barracks, Melbourne
Director General - Army Development (NSO) (4 copies)

Aeronautical Research Laboratories

Director
Library
Superintendent - Structures
Divisional File - Structures
Author: A.S. Machin
J.Y. Mann
J.M. Finney
G.W. Revill
P.W. Beaver
F.G. Harris

Materials Research Laboratories

Director/Library

Defence Research Centre

Library

Air Force Office

Air Force Scientific Adviser
Technical Division Library

Department of Defence Support

Government Aircraft Factories

Library

Department of Aviation

Library

..../cont.

DISTRIBUTION (CONT.)

Statutory and State Authorities and Industry

CSIRO

Materials Science Division, Library

SPARES (10 copies)

TOTAL (53 copies)

Department of Defence

DOCUMENT CONTROL DATA

1. a. AR No	1. b. Establishment No	2. Document Date	3. Task No
AR-002-997	ARL-STRUC-TM-371	January 1984	DST 82/008

4. Title	5. Security a. document	6. No Pages
HOW MANY SPECIMENS? AN AID IN THE DESIGN OF FATIGUE TEST PROGRAMS.	UNCLASSIFIED	9
	b. title: U c. abstract: U	7. No Refs: 5

8. Author(s)	9. Downgrading Instructions
A.S. MACHIN	—

10. Corporate Author and Address	11. Authority (as appropriate) a. Sponsor b. Security c. Downgrading d. Approval
Aeronautical Research Laboratories, P.O. Box 4331, MELBOURNE, VIC. 3001	—

12. Secondary Distribution (of this document)

Approved for Public Release.

Overseas enquirers outside stated limitations should be referred through ASDIS, Defence Information Services Branch, Department of Defence, Campbell Park, CANBERRA ACT 2601

13. a. This document may be ANNOUNCED in catalogues and awareness services available to ...

No limitations.

13. b. Citation for other purposes (ie casual announcement) may be (select) unrestricted (or) as for 13 a.

14. Descriptors		15. COSATI Group
Fatigue tests	Confidence limits	1113
Statistical samples	Experimental design	
Statistical tests		
Specimens (statistics)		
Statistical analysis		

16. Abstract

Increasing fatigue test sample sizes increases the accuracy with which population statistics may be calculated and increases the sensitivity of tests comparing different groups of results, but it also increases testing costs. Information is presented which will aid the designer of a fatigue test program in deciding sample sizes.

This page is to be used to record information which is required by the Establishment for its own use but which will not be added to the DISTIS data base unless specifically requested.

16 Abstract *(Contd)*		
17. Imprint Aeronautical Research Laboratories, Melbourne.		
18 Document Series and Number Structures Technical Memorandum 371	19. Cost Code 25 7020	20. Type of Report and Period Covered
21. Computer Programs Used		
22. Establishment File Ref(s)		